"Spatially Distributed and Stabilized 3-D Surround Sound using Multi-resolution Signal Processing Algorithms for Virtual Environments"

Final Report

Thomas P. Caudell
Department of Electrical and Computer Engineering
University of New Mexico
Albuquerque, NM 87131

ABSTRACT

The Homunculus Project at the University of New Mexico is developing a high performance computing, multidimensional virtual reality laboratory for construction, simulation, evaluation, perception, and comprehension of complex software systems and simulations. This virtual reality interface provides a laboratory without walls in which scientists enter and interact with their software systems, effectively becoming the "little person in the brain of the machine". Capabilities include tools for multidimensional visualization of data flow graphs and visual programs, sampling/displaying of information at random points in graphs for diagnostics, reconfiguration of software modules, and monitoring high performance computers performing the actual calculations. Visual programming graphs are represented in the Homunculus as dynamic, continuously varying conceptual resolution objects joined by multi-resolution information channels. The operator uses an immersive virtual reality head-mounted display to view the living structures and use virtual tools to locomote and navigate through the world, interact with the simulations, and monitor the behavior of the external physical systems under software control. Research continues on the development of infrastructure for this virtual laboratory and the measurement of its effectiveness.

Three dimensional spatially stabilized sound plays a significant role in the representation of information and in the creation of presence in virtual environments. In the Homunculus Project, sound will code operational characteristics of each node in the visual programming graph. For complex graphs, the number of independently stabilized sound sources will exceed today's sound localization technology. To address this problem, we have designed new signal processing algorithms to optimize the use of 3D sound hardware using multi-resolution level-of-detail techniques. This proposal sought funds to purchase new hardware to enable us to test these algorithms and calibrate them with human perception experiments. The primary outcome of this research was to be new software tools that permit virtual environment developers to create perceptually realistic, cost effective, spatially distributed, and stabilized 3D surround sounds with limited hardware resources.

The High Performance Computing and Education Research Center (HPCERC), which operates the Albuquerque Resource Center (ARC) on the campus of the University of New

REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-00-

0342

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** 28 June 2000	**3.** Final - 1 April 1997 - 30 April 1998

4. TITLE AND SUBTITLE
Spatially Distributed and Stabilized 3-D Surround Sound Using Multi-Resolution Signal Processing Algorithms for Virtual Environments

5. FUNDING NUMBERS
F49620-97-1-0144

6. AUTHOR(S)
Mr. Thomas P. Caudell
Department of Electrical and Computer Engineering

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of New Mexico
Albuquerque, NM 87131-3106

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NL
801 North Randolph Street
Arlington, VA 22203-1977

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The Homunculus Project at the University of New Mexico is developing a high performance computing, multidimensional virtual reality laboratory for construction, simulation, evaluation, perception, and comprehension of complex software systems and simulations. This virtual reality interface provides a laboratory without walls in which scientists enter and interact with their software systems, effectively becoming the "little person in the brain of the machine". Capabilities include tools for multidimensional visualization of data flow graphs and visual programs, sammpling/displaying of information at random points in graphs for diagnostics, reconfiguration of software modules,and monitoring high performance computers performing the actual calculations.Visual programming graphs are represented in the Homunculus as dynamic, continuosly varying conceptual resolution objects joined by multi-resolution information channels. The operator uses an immersive virtual realityhead-mounted display to view the living structures and use virtual tools to locomote and navigate through the world, interact with the simulations, and monitor the behavior of the external physical systems under software control.Research continues on the development of infrastructure for this virtual laboratory and the measurement of its effectiveness.

14. SUBJECT TERMS
Multiresolution, algorithms

15. NUMBER OF PAGES
12

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclas	Unclas	Unclas	

Mexico strongly supported this project. The ARC cost shared with this research project by providing and maintaining the primary computing hardware, including a 32 node IBM PowerParallel SP1 Supercomputer and a four processor SGI Onyx with an RE2 graphics engine. and two Research Assistants employed to conduct portions of this research.

1.0 Background of the Homunculus Project

In general, investigating and understanding the content of multidimensional information and its generation process is becoming the major impediment in the application of basic computational research today. For example, massively parallel processors running intricate, complex software systems are being used in a growing number of critical missions in the Department of Energy and the Department of Defense. With conventional human-computer interfaces, the scientist remains separated from his/her software and data, with the computer acting as a recalcitrant intermediary. What is required of human-computer interface development is an approach that allows humans and machines to do what they each do best. We believe that future developments in experimental and computational sciences will critically depend on the development of more effective human-computer interfaces that are designed to use significantly more of our natural human reasoning and perception in the analysis process.

This proposal sought support, through the addition of 3-D sound localization instrumentation, for further development and testing of the Homunculus immersive virtual reality interface to complex software systems and simulations. The use of virtual reality technology provides an opportunity, for the first time in the history of computation, to immerse scientists, with all of their natural abilities in perception and reasoning, directly into multidimensional representations of their data and software.

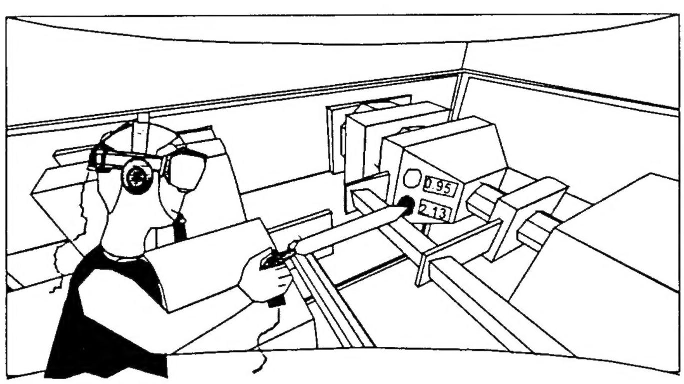

Figure 1. The concept of a scientist immersed in a graphical representation of a complex software system using a hand held tool such as a wand to adjust software parameters.

The current focus of our simulation research is a robotics system, controlled by a complex artificial neural network called the Encephalon. We postulate that the Homunculus environment will act as a virtual laboratory that will expedite research in the field of autonomous perceptual systems such as the Encephalon, as well as other classes of complex software systems, while simultaneously exposing many new research issues in the field of human-computer interfaces.

With virtual reality (VR) technology, the scientist can immerse his/her senses into a virtual environment that contains a representation of complex software systems. When the VR system is interfaced to an autonomous robot, the environment can be thought of metaphorically as an

the sounds in 3-D space, distorting the sounds appropriately for the modeled virtual environment, and playing the sounds for the human listener. In an ideal system, all virtual sound processing would be accomplished with maximum fidelity, minimum cost and in real-time. Real-time for sound processing can be defined as occurring within one frame of video, or less than one-thirtieth of a second. Achieving the ideal system is not possible today due to the many unresolved research questions.

As in the real-world sound process, the first component of the virtual sound system is the sound source. Several components create the virtual sound source: host processor system, signal generator, and signal processor. The target application will largely determine the complexity and functionality of the sound source components. The host processing system controls the VR simulation, the sound signal equipment, and may perform some signal processing functions. The signal generator can include a synthesizer, a sound sampler, a tone generator or simply a large bank of memory for storage of pre-digitized sound samples. Situations where the sounds are not known a priori require dynamic sound generation. This is an active area of research. Today, obtaining even rudimentary dynamic sounds requires extensive equipment and software. Applications with a fixed set of known sounds typically use pre-digitized sounds and thus require minimal signal processing but vast amounts memory.

The signal processing element refers to the hardware and software required for 3-D sound localization, sound mixing and sound source transformations. The 3-D sound system individually localizes each sound source according to the user's current head position. Head position information is obtained via a tracker placed on the user's head. The amount of signal processing equipment required increases as the number of participants and sound sources increase.

The environmental processing component refers to both the modeling of the virtual world and overcoming distortion effects of the participant's physical environment. Accurate modeling of a virtual environment requires significant signal processing equipment. Environmental modeling is important due to the vast amounts of information humans obtain from environmental distortions. Development of techniques for simplifying environmental distortion modeling are on-going. It is possible, but difficult, to overcome sound wave distortions created by the participant's physical environment. Headphones are an easy way to minimize the distance between the sound source and the receiver and thus minimize the environmental distortion. If the sound is displayed using speakers, real-world environmental distortions (which typically do not correlate with the virtual environment) will modify the sound. The types of headphones that fit inside the ear are often preferred for 3-D sound systems because they minimize environmental distortions, have lower resonance than larger headphones that enclose the outer ear, and provide relatively good attenuation of external sounds [Durlach 95]. Active noise cancellation equipment can serve to further control the sound environment.

The last link in the virtual sound process is the human listener. Luckily, the human auditory system need not be simulated or modeled. However, determining the effectiveness of the simulated sounds requires at least a rudimentary understanding of human sound perception. This is being accomplished through extensive psycho-acoustic experiments. Significant on-going research continues in this area as the perceived importance of the auditory sensory system increases.

2.2 THREE-DIMENSIONAL SOUND

Several different terms are synonymous with 3-D sound: sound localization, spatialized sound, binaural audio or virtual acoustics. There are several compelling reasons for using 3-D sound. As discussed earlier, the use of 3-D sound is useful for discriminating between sounds and in directing the human's attention towards an urgent sound. Humans hear sounds spatially in the real world, thus in creating realistic virtual environment simulations, sounds should be heard in the same way. This is especially important when simulating high-stress scenarios because often times the high stress is due to the intense, encompassing sound environment.

Over the past several years, techniques have been developed which allow sound placement in a 3-D environment around a human listener. Some sound localization systems on the market today use only time difference and intensity differences to locate a sound within a virtual environment without taking into consideration the distortions created due to the head, torso, and outer ear of the listener. As a result, these systems are lacking in horizontal direction accuracy, accurate vertical location ability and the externalization (out-of-the-head) sensation. Systems that additionally include digital filters to model the head, torso, and outer ear distortions are much more effective in achieving 3-D sound simulation. These filters are often referred to as head-related transfer functions (HRTFs). By filtering a digitized sound source with the appropriate HRTF filter for the user's current head position (obtained from head position/tracker sensors), one can potentially place sounds anywhere in the virtual space about a listener. These HRTFs are used as the basis of the digital filters for processing the synthesized sound. For a more detailed description of 3-D sound and the HRTF approach see [Wenzel 92] [Begault 94]. HRTFs vary somewhat between listeners, because the shape and size of each person's head, torso and outer ear are unique. If a generic or standard HRTF is used, the variation is often times sufficient to create significant errors in perceived sound localization.

3.0 Summary of Research that Required Instrumentation

The cost of a virtual sound system relates directly to the system's performance specification. If the application requires high speed and high quality sound rendering and localization, the equipment can be very expensive. For example, today sound mixing boards range from $100 to more than $10,000 for professional sound mixers. As of this writing, sound localization systems range from $1,500 to $10,000 per sound channel depending on environmental modeling complexity. With current techniques, sound sources and individual reflections require separate sound channels; thus, the localization system cost increases directly with each modeled source or reflection.

Ideally, the designer of a virtual environment would be able, as the application demands, to place an arbitrary number of sound point sources in the three dimensional volume accessible to the user. To produce true arbitrary 3D surround sound, the localization system requires a sound channel for every "sound pixel" perceivable by a human user. Using localization error measurements as an estimator of the angular size of such a pixel, we can calculate the number of required channels: $N = 4 / \tan^2(\theta_{error})$. For an error of ±5 degrees there are 522 sound pixels in a sphere; for an error of ±1 degree the number rockets up to 13,128 sound pixels. At a price of $2,000 per channel, the cost of a surround system would be between $1 million and $26 million, clearly a ridiculous price.

To address this problem, we propose to study the use of "level-of-detail" techniques for sound localization in complex sound environments and for spatially distributed sound sources. In graphics, simpler polygon models are used for objects distant from the user, while more complex ones are used for near objects. The polygonal complexity is changes as the distances change between object and user. This approach can improve rendering performance for graphics systems while minimally impacting the perceived visual effects. For 3-D sound, rendering is replaced with signal processing. In the context of sound localization, level-of-detail means using a distance-dependent spatially smeared HRTF in conjunction with weighted combinations of point sound sources, to approximate the perception of a given environment. This is explained in more detail in the next section. These algorithms view the fixed (small) number of convolution channels of the hardware as a computational resource, whose use is optimized based on a yet to be measured perceptual merit function.

3.1 Signal Processing Algorithms

To place a sound in a 3-D environment, the digital filter corresponding to the desired target location is convolved with the sound signal to be localized and current head position information of the user. The convolution process can be performed by multiplying the discrete Fourier transform (DFT) of the original signal with the DFT of the digital filter. For this research, the Aureal Acoustetron II was used as the computational engine. The process of convolution places a sound signal in the perceptual 3-D space of the listener. This technique is most often used with headphones, rather than speakers, in order to minimize environmental effects. System processing capabilities and high per-channel cost typically limits the number of simulation channels in a system.

The Aureal Acoustetron II sound localization hardware uses digital signal processing chips to perform the real-time convolution of an input point sound signal with an HRTF. The HRTF is selected for its characterization of the azimuth(θ) and elevation(ϕ) of the signals origin relative to the current head position and orientation. Stated in the form of equations,

$$P[i] = S[\theta_i, \phi_i] * H[\theta_i, \phi_i]$$

where $P[i]$ is the i^{th} perceived spatially localized sound entering the ear canal, S is the point sound signal located in direction $[\theta_i, \phi_i]$ in head coordinates, H is the HRTF for that direction and ear (left or right), and * represents the convolution integral transform operator. The total experience is the sum over sound signals,

$$P = \sum P[i] = \sum S[\theta_i, \phi_i] * H[\theta_i, \phi_i], \quad 1 \leq i \leq N$$

where P is the total sound entering the ear canal, and N is the total number of point sound signals. Convolution is a linear process. Therefore, if a constellation of point sound sources all originate in the same general direction of space, the summation in the above equation may be moved inside the * operator,

$$P = S[\theta_c, \phi_c] * \sum H[\theta_i, \phi_i] = S[\theta_c, \phi_c] * H[\theta_c, \phi_c],$$

where $S[\theta_c, \phi_c]$ is the common sound for the region and $H[\theta_c, \phi_c]$ is the effective smeared HRTF for the common region of space. In this way, N convolution operations have been reduced to one, requiring only one channel of hardware. The resulting sound would be perceived as emanating from a common angular region of space. Again using the linearity of the convolution operator, if we are far enough away from a collection of different point sound sources, we may approximate the sound perception at the listener by performing a weighted average of the sounds,

$$P = \sum w[\theta_i, \phi_i]\, S[\theta_i, \phi_i] \;*\; \sum w[\theta_i, \phi_i]\, H[\theta_i, \phi_i] = S[\theta_w, \phi_w] * H[\theta_w, \phi_w],$$

where $w[\theta_i, \phi_i]$ is a weighting function for the i^{th} sound signal, $S[\theta_w, \phi_w]$ is the effective signal for the region, and $H[\theta_w, \phi_w]$ is the effective smeared HRTF for the common region of space. Again, we have reduced N convolutions to one with this approximation.

These are examples of signal processing techniques that can be applied to reduce the amount of computation needed to approximate a true surround sound environment in real-time. The connection to multi-resolution level of detail can be seen in the last example. As the user moves about the virtual environment, the resolution of the sound environment will change to keep the amount of computation within the bounds of the existing localization hardware. The criterion for when and how to make this change must be determined through human studies.

In order to implement and test these concepts, the level of detail algorithms must have access to the HRTF data and the real-time signal processing algorithms. The following section summarizes the goals of this project and gives status on the progress (and obstacles) towards them.

4.0 Results of Research

For complex virtual environment, the number of independently spatially stabilized sound sources will exceed today's sound localization technology, including a single Acoustetron processor. To address this problem, we have presented new signal processing algorithms to optimize the use of the 3D sound hardware using multi-resolution level-of-detail techniques. The funds under this grant were used to purchase new sound hardware to enable us to test these algorithms and calibrate them with human perception experiments. The primary outcome of this research was to be new software tools that permit virtual environment developers to create perceptually realistic, cost effective, spatially distributed, and stabilized 3D surround sounds with limited hardware resources.

A client-server software architecture was used for this research. The server code executed on the array of Acoustetron processors, while the client code ran on an SGI O2 graphics workstation. When the virtual world required sounds or changes in existing sounds, a command was sent over the local Ethernet to the Acoustetrons. The resulting sounds were channeled through a audio mixer and ultimately to either speakers or earphones. All sounds for this work are stored as standard Wave files. Therefore the practical goals of this research were to:

1) Integrate an array of three Acoustetrons into one virtual localization resource and establish a connection to the SGI processor

2) Implement the Level-of-Detail algorithms and test

3) Experiment with the resulting system to measure human perceptual thresholds.

This section discusses the technical progress in detail vis-à-vis these three goals.

4.1 GOAL 1: INTEGRATION

The first goal was accomplished early on - software was written that provides a reasonable approximation of a seamless virtual resource. Issues were uncovered that prevent a complete integration into a seamless resource, mostly due to limitations inherent in the Acoustetron server software:

a) Waveforms cannot be streamed to the Acoustetrons over the Ethernet - they must be uploaded, saved to disk, and then used. This is a large barrier to real-time synthesis, and also to virtualization.

b) Acoustetrons have audio distortions under load, especially at startup, that manifest mostly as phase and timebase/frequency distortion. This is apparently caused by the server software as it streams digital data from the hard drive to memory and thence out the 3D sound cards. It is especially evident at startup as a "pitch bend", and in operation it manifests itself as a timebase drift, such that signals previously in synchronization drift apart. For an assortment of varied signals, this is acceptable, but it is mostly evident when playing multiple copies of the same waveform.

The software can handle a large number of Acoustetron servers, from one to a few dozen, with very little load. The code utilizes a few O(n) algorithms, but in general the software could easily be tuned to serve large numbers of hosts, particularly if the network were upgraded to, say, fast Ethernet.

The current software is a Unix-based library that presents an API similar to the provided Acoustetron API, with the following extensions:

a) One can specify which machine and input to use for a particular sound, in the interests of attaching an analog input from, for example, a speech synthesis system.

b) One can also specify which machine, leaving the channel choice up to the software, for the case when an assortment of sounds are stored on a particular system.

c) If the user leaves the choice up the software, it attempt to distribute channels across the available Acoustetrons, approximating static load-balancing.

d) When using stored Wave files, the user must specify which machine the sound is stored on. This prevents duplication of large data files.

Currently, we have several demonstration programs that exercise the library, Acoustetrons, and communications network, and have successfully integrated the library into our virtual environment. (MuSE) We are now planning how to extend the sound capabilities of MuSE

further, via external systems such as the Acoustetrons, or solutions based on software rendering. The current system can localize 24 stored 44 kHz waveforms, or twice that many at 22 kHz. If one uses analog inputs, the system can localize 24 of sources at any sampling rate up to 44 kHz, provided that all inputs are sampled at the same rate.

4.2 GOAL 2: LEVEL-OF-DETAIL ALGORITHMS

This goal has not been accomplished, requiring some explanation. Out the outset of this project, we contacted the Acoustetron manufacturer and formed and agreement that would have given us access to the Acoustetron server software and its HRTF tables. As mentioned before, this information is required to implement the level-of-detail algorithms. Specifically the algorithms would be modifying neighboring HRTF table entries on the "fly", substituting the calculated low spatial resolution HRTFs. However, access to the internal workings of the code required a nondisclosure agreement with Aureal that proved impossible to obtain, even after may promises. Another challenge to this goal was that of support. During the time that we they were making commitments to support this research through access to the software, and processing our purchase orders for the equipment, Aureal was quietly dropping the Acoustetron from their product line. This clearly limited the level of technical assistance we could expect when attempting to work with their software and data sets.

Given these problems, we decided to decline signing the NDA, and have pursued other avenues for dealing with localization.

4.3 GOAL 3: EXPERIMENTATION

A series of perceptual experiments were planned to validate the realistic quality of the localized sounds. The first experiment would have involved an acoustic quality rating test, where a fixed set of sounds will be progressively averaged using the algorithm outlined in the previous section. This would have provided a measure of "goodness" for the localization of sounds as a function of level-of-detail. These experiments will require a large number of localization channels to simulate a "high" resolution sound source to be compared to a series of reduced resolution representations. For example, an 8 x 8 grid of high resolution sounds will need to be produced in front of a subject playing 64 identical pre-digitized sound signals. The subject was to be asked to detect when the "character" of the sound changes as the resolution is reduced using the algorithms outlined in the previous section.

Other experiment were planned to evaluate the perceptual quality of the localized sound, but given the lack of cooperation with the vender of our 3D localization equipment, none of these experiments were performed. We are making other plans to continue this aspect of the research, as will be discussed next.

5.0 Future Directions

The question that naturally arises in these discussions is that of scale - to wit, how many sounds can a given system produce at once? We believe that compelling virtual audio environments will require hundreds to thousands of such sounds, and wish to pursue avenues that

will produce such a system. Lack of cooperation not withstanding, the Acoustetrons can each handle eight signals and are therefore not reasonable for such a large-scale systems.

Other portions of our research program are using wavelets for sound representation and synthesis. We have conducted research involving wavelets and perceptual effects for environmental synthesis, which is of use for background or environmental audio such as rainfall, wind noise, and so forth. To complement this work, we are studying the use of wavelets in the localization process, with the goal of building a system that uses wavelet-based filter banks for localization for the following reasons:

a) Wavelets provide nearly optimal tools for experimenting with HRTF compression via coefficient truncation and basis selection. This has the potential to greatly reduce the memory required in storing the HRTF tables.

b) A wavelet-based system would avoid two or more Fast Fourier transforms in the localization process, and could therefore be made more computationally efficient. This implies better scalability.

c) If we can localize in the wavelet domain, we could easily integrate wavelet synthesis tools, for an efficient method of producing dynamic environmental audio. Additionally, this would reduce the memory required, since we need only the baseline sample and the coefficient modification algorithm.

We are also researching other methods of large-scale sound production, using multiple computers or CPUs, and performing the localization algorithm in parallel to achieve many simultaneous sounds. We are also studying the use of off-the-shelf hardware such as embedded processors, CPLDs, DPSs, or FPGAs, to build a large scale system that performs the localization parallel using a logarithmic adder tree.

All of these approaches have complex tradeoffs in cost, complexity, flexibility, latency, and scalability and are also amenable to using either Fourier or wavelet-based localization.

5.1 Software Wavelet Localization

Many current systems such as SGI Onyxes have multiple CPUs. If we can run jobs on more than one processor, we can run several copies of the localization algorithm, and thereby increase the number of feasible sounds. A variant of this is to use an external parallel system, such as an IBM SP2, as a sound localization array. This provides the benefits of commodity parallel hardware, and hopefully the cost benefits thereof as well. However, there are some associated issues that must be resolved:

a) Mixing of the output.

Do we have a server that does an arithmetic mix of the output streams? This requires a communication channel with large bandwidth and deterministic latency, but is noise-free and feasible using primitives such as MPI AllGather. We could also place analog output hardware in each node (commodity sound cards for PCs cost less than forty dollars at present) and mix the

result using analog electronics. This does not require the scalable communications channel of the digital mixing, but can introduce noise and requires a mixing board or equivalent to sum the outputs. Also, some hardware (i.e. the aforementioned SP2) does not have sound output, and is costly to obtain.

b) Latency

We have obtained results on the acceptable latency in generating and localizing audio, so we know that we have a latency budget of 100ms from the VR event to the audio output. We must design our system to accommodate this.

c) Comminations cost

For some sounds, such as events, we will need only communicate a small message from the VR system to the rendering host. If we wish a large number of rendering hosts, the channel must have sufficient capacity to handle the message without exacerbating the latency problem. As noted above, digital mixing and dynamic synthesis both require large bandwidth which also be considered.

d) Complexity

While each rendering host runs essentially the same software, there is considerable complexity in the algorithms required for communications, rendering and mixing. Debugging such a parallel, real-time system would undoubtedly be quite challenging.

5.2 Dedicated Hardware

Using off-the-shelf commercial hardware such as PC/104 embedded processors, one could build a system that does localization in parallel on general or special purpose hardware. There is a spectrum of possible approaches, ranging from VHDL-programmed CPLDs to C code running on a DSP or microprocessor node. The results can be mixed onboard, using an adder tree or analog mixer. This approach also has its pros and cons:

a) Opacity/complexity

Special purpose hardware is more difficult to design and debug than software on a desktop computer. One requires tools such as design software, logic analyzers, and oscilloscopes whose cost must be accounted for in the design selection process.

b) Communications cost

This approach solves the communications problem, since the mixing is done onboard by a scalable adder tree.

c) Scalability

Each node could run one or more copies of the localization algorithm, depending on how much computation power each has. For example, Texas Instruments has DSPs capable of over a

billion operations per second; one could expect to run quite a few copies on such a node. Additionally, the logarithmic adder tree offers excellent scalability, though one cannot easily expand it once built.

d) Latency

This is less of an issue, since the mixing is greatly accelerated. The main concern becomes ensuring that each node can localize within the allotted time budget.

6.0 Conclusion

Three dimensional spatially stabilized sound will play a significant role in the creation of presence in virtual environments, as well as the representation of information. In the Homunculus Project, sound will code operational characteristics of each node in the visual programming graph. For complex graphs, the number of independently stabilized sound sources will exceed the capability of today's technology. To address this problem, we have designed a new set of signal processing algorithms to optimize the use of 3-D sound hardware using multi-resolution level of detail techniques. This report discussed the purchasing of new hardware to enable us to test these algorithms and calibrate them by human perception experiments. This research will allow virtual environment developers to create perceptually realistic spatially distributed and stabilized 3-D surround sound with limited hardware resources.